Modelling Aspects of Reduction and Assimilation of consonant sequences in Spontaneous French Speech

Danielle Duez

MODELLING ASPECTS OF REDUCTION AND ASSIMILATION OF CONSONANT SEQUENCES IN SPONTANEOUS FRENCH SPEECH

Danielle Duez

Laboratoire Parole et Langage, CNRS UMR 6057

ABSTRACT

The following paper first presents spectrographic data of consonant sequences containing one or two consonants omitted and/or changed into another consonant when compared to a previous perception analysis. In most cases, perceptual and acoustic data are seen to strongly correspond, proving that consonants had indeed be changed, significantly reduced or deleted, mainly in a weak position, thereby preserving acoustic information crucial for lexical access, integration of prosodic structure and successful communication. Finally, tentative rules summarise the tendencies observed in reduction and assimilation patterns.

1. INTRODUCTION

The present study examines some of the effects of reduction and contextual assimilation on the acoustic structure of two-consonant sequences in spontaneous French speech, both consonants occurring initially or finally (/C1C2/) in a syllable and consonants occurring in two consecutive syllables (/C1#C2/ and /C1(ə)C1/).

In his hypo/hyperarticulation model, [1] described natural speech production as an adaptive process by which speakers adjust their performance according to communication goals and situation demands. In casual speech styles such as conversations, speakers speak fast and hypoarticulate, decreasing the duration and amplitude of phonetic gestures and increasing their temporal overlap. At the acoustic level, hypoarticulation is reflected by greater undershoot [2] and contextual assimilation of speech segments, with words being reduced, omitted, or combined with other segments compared to the phonetic segments of the same words when read.

Hypoarticulation effects are not uniform. Fundamentally, hypoarticulation is governed by a speaker's need to produce an auditory signal which possesses sufficient discriminatory power for successful word recognition and communication [1]. The investigation of reduction and contextual assimilation processes in conversational speech allows a better understanding of basic principles governing them and provides an explanation of their occurrence and in what manner. In this respect, the analysis of consonant sequences in conversational speech is of special interest since consonants in sequences are not discretely produced and are submitted to a wide range of interacting linguistic influences.

Section 1 summarises spectrographic data of consonant sequences containing one or two consonants reported as omitted and/or changed into another consonant when compared to a previous perceptual investigation [3]. Consonant omission is the result of an extreme reduction process while consonant change reflects a total assimilation. Spectrographic data were compared with perceptual data and underlying phonological sequences, revealing strong links between perceived consonant and immediately adjacent consonant and vowels. Section 2 reviews the interaction of reduction and assimilatory processes by type of word (function or content), position in syllables and words (initial, medial or final), and position in phrases (final or non-final). Section 3 proposes a certain number of rules identified when analysing reduction and assimilation, particularly adapted for speech technologies.

2. SPECTROGRAPHIC ANALYSIS

2.1. Perceptual criteria of selected consonant-sequences

In [3], 720 sequences extracted with adjacent vowels from the conversational speech of two male French speakers (Spk1 and Spk2) were presented for identification to 16 listeners, who were asked to write down everything they heard using orthographic transcription. There were two series of 360 $V_1C_1C_2V_2$ sequences (one per speaker), each contained the nine following groups OO, OF, OS, FF, FO, FS, SS, SO, SF, where O is an occlusive, F a fricative and S a sonorant. Each group had 40 C_1C_2 sequences. Each sequence was assigned a single identification value corresponding to the phonological underlying sequence reported by at least 75% of listeners. This yielded two main groups:(1) sequences identified as the underlying phonological sequences and (2) sequences identified as different. In the latter group, there were four cases:(2a) the phonological C1 and/or C2 was identified as another consonant (C), (2b) one or two features were identified, (2c) neither a C nor a feature was identified, and (2d) the C was omitted. The number of sequences different from phonological sequences is high (Spk1: 179, Spk2: 166). It is mainly correlated to the high number of sequences with a C1 different from phonological (Spk1: 126, Spk2: 109), the corresponding numbers for C2's are 28 and 28, for sequences with both different C1's and C2's, 26 (Spk1) and29 (Spk2).

To focus on extreme cases of reduction and assimilation, the current presentation is limited to omitted C's and to C's changed from one to another. Their repartition is as follows: sequences with an omitted C1 (Spk1: 38; Spk2: 31) or an omitted C2 (Spk1: 13, Spk2: 2), sequences with another C1 (spk1: 18, Spk2: 9) or another C2 (Spk1: 2, Spk2: 4), and sequences with both an omitted or changed C1 or C2 associated to a changed, omitted or unindentified C1 or C2 (Spk1: 7, Spk2: 14).

<u>Different C1's.</u> For Spk1, there were 5 F's (3/ʒ/'s, 2 /s/'s, 1 /v/) and 3 O's (2/d/'s and 1/p/) reported as voiced or devoiced. For Spk2, the corresponding numbers were 5 F's (2/s/'s , 1 /ʒ/, 1 /v/, and 1/z/), and two O's (2/d/'s). Some oral O's were reported as their nasal counterparts (Spk1: 5 /d/'s, 1 /t/ and 1 /b/; Spk2: 1 /t/

and 1 /d/). One /n/ followed by a /p/ was reported as /m/ (Spk1).

Omitted C1's. Reduction processes affected C1's in the following order: 1) S's (Spk1: 27; Spk2: 20), O's (Spk1:5 and Spk2: 15), and 3) F's (Spk1: 2; Spk2: 8).

Different C2's. One /d / preceded by a /z/ was reported as /v/ and one /p/ followed by /ɔ̃/ as /m/ (Spk1); One /p/ preceded by an /s/ was reported as /f/ and one /v/ preceded by an /s/ as /f/ (Spk2).

Omitted C2's. The distribution was as follows: S's (Spk1: 4, Spk2: 2), voiced (Spk1: 4) or unvoiced O's (Spk1: 2), and voiced F's (Spk1: 2).

Different C1's+different, omitted or unidentified C2's. There were 4 cases for Spk1 and 11 cases for Spk2

Omitted C1's+different, or unidentified C2's: There were 5 and 3 cases for Spk1 and Spk2, respectively.

2.2. Method

2.2.1. C-segmentation

Spectrograms and oscillograms were generated for each sequence. Sequence onset and offset coincide with the preceding-V offset and following-V onset, respectively. The boundary between C1 and C2 was labeled as follows:

O+(O, F or S): the boundary occurs at the end of the burst of the first O (voiced or unvoiced). In case of absence of a burst, the discontinuity (if visible) between the two occlusions (O+O), between occlusion end and noise beginning (O+F), or the beginning of noise or F*2* (O+S) was the boundary.

F+(F, O or S): F differences mainly rely on differences in intensity and noise zone. Therefore, the F/F-boundary location was based on changes in spectral zone, discontinuity in voicing, and change in intensity. The complete cessation of the F noise coincided with the beginning of S (F*2* or noise).

S+(O,F or S). The cessation of the S-F2 coincided with the appearance of voice bar or silence (S+O), of frication noise (S+F) or the change in harmonic density, spectral repartition of formants or formantic structure (S+S).

2.2.2. Acoustic-cue analysis

Acoustic patterns of C's were examined in relation to perceptual data and phonological C's. The analysis was restricted to cues proven to be relevant in C identification.

O's [p,t, k, b, d, g]. For each phonological and perceived O, the presence of an occlusion and/or a burst was controlled, the duration of visible occlusions and bursts measured. Burst frequency, known to be a place cue, was also measured with LPC. Voiced O's have low frequencies, their presence was checked and duration measured.

F's [f, s, ʃ, v, z, ʒ]. The presence of noise, which is typical of F's [4] was controlled. The limits and maximum of noise frequency, which are place cues, were measured. Voiced F's may have two or three formants, depending on the proportion of voicing and fricative elements. Voiced-F patterns were examined and fricative duration measured.

S's [m, n, l, ʀ]. S's have a formantic structure. The spectrum of nasals is dominated by a strong low-frequency formant [4] and characterized by mid-frequency formants [5]. Laterals are characterized by stronger mid-frequency formants than nasals. The French dorsovelar /ʀ/ is context-dependent : devoiced in an unvoiced context, vocalic in a voiced context.

2.2.3. Analysis of reduction and assimilation

For oral O's reported as nasals, the presence of mid-frequency formants was examined. Three patterns of nasalisation (N) were determined: no N (no formants), partial N (a separate stop exhibiting an occlusion and/or burst) and total N (no interruption of mid-frequency formants). Similarly, three patterns of voicing or devoicing were defined for O's and F's, depending on the presence of low frequencies. The same procedure was used to investigate frication of O's. For place change of O's and F's, we measured burst frequency, noise limits and maximum, for nasals place change in V/C transitions. The absence of noise, closure and burst, and formants was controlled for omitted F's, O's and S's, respectively. Present segments (if any) were measured using the above procedure. To check whether for a trace of underlying Cs, the duration of remaining C1's and C2's measured and compared with the duration of intervocalic reference C's.

2.2.4. Syllabic structure, word type and position in syllables and phrases

Sequences were divided into homosyllabic [#C_1C_2 or C_1C_2#] and heterosyllabic sequences [C_1 #C_2] or C1#C1]. In [C_1 #C_2]'s, C1 and C2 are coda and onset C, respectively. [C1#C1]'s result from the deletion of an optional mute /ɔ/ within a word or in a function word [C1(ə)#C1], each C is a C1. C's were also analysed as a function of location in function words (articles, pronouns..) or content words (nouns, verbs). With no lexical stress in French, prominent syllables are mostly phrase-final syllables to which prominence was assigned. A prominent syllable was defined as a syllable which stands out from others in its environment (i.e.longer, more intense, and/or higher in pitch)

2.3 Results

2.3.1. Oral O's reported as nasals

C1's. The 6/d/'s and the 2/t/'s perceived as /n/ in nasal vowel context have complete overlapping of mid-frequency formants with occlusion. All have at least two visible formants (about 1200 –Hz, 2000 Hz or 2700 hz and 4000 Hz). For Spk1, the /d/'s are 45 ms, 57 ms, 64 ms and 77 ms long, the /t/ is 94 ms long. For Spk2, the /d/ and the /t/ are 67 ms and 55 ms long. None has a burst. Surprisnigly, 2 /b/'s and 1 /d/ were reported as /m/ and /n/ in an oral vowel context (Spk1 and Spk2): they may be confusions or misproductipons. One /m/ has mid-frequency formants (1100 Hz and 2600 Hz) and is 67 ms long ; the other has an occlusion (70 ms) and a burst (7ms, 1700 Hz). No mid-frequency formants and no voice bar for the /n/. Transitions may allow listeners to identify articulation place.

C2's. A /p/ followed by a nasal vowel (/ɔ̃/) is reported as /m/: it is 74 ms long and has a 1000-Hz formant. Again, a /t/ reported as /n/ in an oral vowel context there is no trace of mid-frequency formants and burst.

2.3.2. Voiced or devoiced O's and F's

C1's. No low frequencies for 4/ʒ/'s, 1/v/ and 1/z/ and 4/d/'s reported as unvoiced; on the contrary, low frequencies for 4/s/, 1/p/ and 1/t/ reported as voiced. For all them, there was anticipatory effect of a unvoiced or voiced C2. Although reported as unvoiced 1/ʒ/ is partially voiced (25 ms out of 50ms), 1/ʒ/, 1/v/

and 1/d/ have low frequencies.
C2's. There are no low frequencies, but a short frication noise (51ms) for a /v/ reported as /f/ in a /sv/ sequence.
C1+C2's. A /ts/ sequence is totally voiced: both C's have low frequencies. On the opposite, a voiced sequence (/dʒ/) reported as unvoiced has no voice bar. Six single C's result from the coalescence of C1-manner-and-place and C_2 voicing: /sd/=>/z/, /sv/=>/z/, /fz/=>v/, /ʒs/=/ʃ/, /tv/=>/d/. The /z/'s and the /v/ have low frequencies, their respective duration is 150 ms, 120 ms, and 94 ms. The /d/ has both a voice bar and burst (3750 Hz), its duration is 120 ms. The two /ʃ/'s have no low frequencies, their duration is 100ms and 175 ms, respectively. Each reported C is longer than the mean duration of intervocalic /z/'s (70ms, SD: 20), /d/'s (74ms, SD: 34), /ʃ/ (112ms) and /v/'s (65 ms, SD: 25).

2.3.4. Other cases
C1's. Two /n/'s preceded by a labial vowel were reported as /m/. one 20-ms /m/ has a 2500-Hz formant and flat V/C transitions. The other is 65 ms long, has three visible formants (1117 Hz, 2194 Hz and 4209 Hz) and there is a clear downward movement of the second and third formants at the preceding-V end. A /ʀ/ followed by /l/ was reported as /b/, With no occlusion, burst, and formants, it may be a confusion. There are no visible occlusion, burst, and formants. A /f/ results from /t/-manner-and-place change (/tʀ/ =>/f/), the other from of a /s/-place change (/st/=>/f/). Both exhibit a weak noise whose limit is around 1000 Hz, their duration is 60 ms and 115 ms, respectively.
C2's. A /p/ in /sp/ is changed into an /f/ with a weak frication noise (limit: 500 Hz, duration: 65 ms). A /l/ in /sl/ is changed into a typical /t/ (silence :50ms; burst : 14 ms, frequency: 4200 Hz). A /b/ reported as /g/ has a voiced bar (62ms), and a burst (18ms, frequency: 1669Hz). A 105-ms /d/ changed into /v/ has a 1200-Hz formant. The /t, g and v/ may be errors.
C1 and C2. A /z/ resulting from the merger of /d/ place and /v/ manner has a visible frication noise (limit: 3500 Hz, maximum: 4500 Hz).

2.3.4. *Omitted C1's and/or C2's*
C1's. One /l/ and two /ʀ/'s were omitted despite visible mid-frequency formants (1500Hz; 1472Hz and 2281 Hz, and 2300 Hz). Their short duration (30ms, 18 ms, 30 ms) may have not given listeners sufficient time for identifying them. There is no trace of formants for the other S's which were deleted. The duration of the remaining C2's is highly variable and often equal or inferior to reference C's. Two identical C's usually coalesced into a single C whose duration is often longer than that of reference C's. For example, /ʀ/ duration is 96ms, 76 ms for Spk1 and 77 ms for Spk2: mean reference duration: 58 ms and 52 ms for the two speakers. C1's and C2's having the same place and manner tend to combine into a single C whose duration is also longer than that of reference C's: Spk1 (/s/:156 ms, reference /s/: 104 ms); Spk2 (//f/: 145 ms, reference //f/: 88 ms; /t/: 94 ms and 122 ms, reference /t/: 91 ms). For C1's and C2's with the same manner and voicing, the trend is similar although less accused. The remaining /z/, /p/ and /d/ in /gz/, /tp/ and /kd/ sequences is 100 ms, 114ms and 105 ms, respectively: the mean reference duration of /z/'s, /p/'s and /d/'s is 70 ms, 90 ms and 61ms. In addition, there is often a trace of the underlying C in preceding V transitions: e.g. in case of omitted velar, there is a merger of the second and third formants at the end of the preceding V.
C2's. As for C1's a few S's (1/m and 2 /l/) were reported as omitted in spite of their visible formants. Again their short duration (15ms, 20 ms, respectively) may be a reason to listeners' omission. There are no visible formants for the other omitted sonorants (Spk1: 2 /ʀ/'s and 3/ /l/'s; Spk2: 1 /ʀ/ and 2/ /l/'s. The duration of the remaining C1's does not exceed that of reference C's. Four O's having the same articulation place as their preceding C1's were omitted: 2 /d/'s preceded by a /z/ and a /t/, respectively, and 2 /t's belonging to /st/'s. The remaining /z/ has a duration (75 ms) close to that of reference /z/'s (78 ms), contrary to the two remaining /s/'s which are much longer (145 ms and 175 ms, respectively) than reference /s/'s. The remaining /t/ of /td/ sequence has no voice bar and a duration equal to that of reference C's.

2.3. Discussion

In some cases, listener reports differ from spectral cues, due to confusions or misperceptions induced occasionnaly by the artificial truncation of vowels used in perception tasks. Misproductions are frequent in spontaneous speech, there may be some in the above acoustic data. However, generally, perceptual and acoustic data strongly correspond, proving that C's were changed, significantly reduced or deleted. Two main tendencies emerge for assimilated C's:1) the nasalisation of O's, which can be viewed as the result of overlapping of the velum-lowering gesture of a preceding (or more rarely following) nasal vowel and the closure gesture of an occlusive and 2) the devoicing or voicing of voiced or unvoiced C1's and/or C2's (O's or F's) due to the anticipatory and/or the carryover effect of an unvoiced or voiced C2 or C1. Less frequent were the cases of O's frication, mainly due to the fricative gesture overlapping the closure gesture, and the merger of C1 place and manner with C2 voicing. Strikingly, most changed C's keep their place of articulation, confirming results on intervocalic voiced stops [6]. Omitted S's are very short or deleted, reflecting a strong decrease in the magnitude of the movements associated with the gesture, or the complete deletion of the gesture [7]. Other omissions reflect assimilatory processes leading to the deletion of O's or F's. In some cases, there is trace of some residual gestures in V/C transitions and/or in the long duration of single remaining C's.

3. SYLLABLE STRUCTURE, WORD TYPE, LOCATION IN SYLLABLES AND PHRASES

The repartition of assimilated, omitted, unidentified and identified C's as a function of the four above factors is significant for Spk1 (.0001). Out of the 34 deleted C's, 33 are in heterosyllabic sequences, 30 in coda position, 32 in non-final syllables and 20 in function words. The reduced form of function words such as "il" (he)=>/i/ is very frequent. Out of the 22 assimilated C's, 18 belong to heterosyllabic sequences, 19 to non-final syllables. The effect of position in syllables and word type is weaker than for deleted C's: 10 are in function words, 12 in content ones, 9 were coda C's and 9 C1's in C1#C1 sequences. For Spk2, there is also a significant effect of syllable structure and location in syllables: out of the 43 deleted C's, 36 are in heterosyllabic C's, 31 in coda position and 5 in C1#C1's ; for the 20 assimilated C's the repartition is as follows: 17 in heterosyllabic sequences, 11 in coda position and 6 in C1#C1's. Deleted and assimilated C's are mainly located in non-final syllables (38 and 18, respectively), however this repartition is not significant.

For C2’s, there is a significant effect of location in phrases, but not of location in syllables for Spk1. For Spk2, it is the opposite. There are 8 out of 10 deleted C’s in heterosyllabic sequences, 4 belong to C1#C1 sequences and 4 are onset C’s. Deleted and assimilated C’s are mostly in non-final syllables. Omitted and assimilated C2’s are rare, however they are interesting cases: most of the function words with a dropped /ə/ tend to be deleted, the first C of a verb after the clitic /ʒ/ (je: “I” is strongly reduced or deleted when voiced, /ʒ/, followed by an unvoiced C, is devoiced.

Reduction and assimilatory processes particularly affect coda C’s in non-final syllables, thereby preserving sequences which carry all information crucial for lexical access, integration of prosodic structure and successful communication.

4. TENTATIVE RULES

Spontaneous speech is well known for its extreme variability in speech segments. Even so, the acoustic data obtained reveal certain regularities in patterns of reduction and assimilation. The following tentative rules attempt to summarise some of the aspects of reduction and assimilation.

Each rule is structured X=>Y/W__Z where X is rewritten after left-hand context W and before right-hand context Z. If Y is absent, there is a deletion rule. Each symbol is composed of a phonetic segment (V and C for vowels and C’s, respectively) or boundary name (full stop for syllable boundary). Vowels and C’s are shown as a function of binary features:+/-FUNC means function/non-function word marker. As assimilated, reduced or omitted C’s were mainly in non-prominent syllables, the feature /-PROM/ is not represented. Variables α, β..mark compatibility between the signs of features.

Nasalisation of voiced and unvoiced O’s after a nasal vowel

$$\begin{bmatrix} C \\ -nas \\ +occ \end{bmatrix} \rightarrow [+\text{nas}] \;/\; \begin{bmatrix} V \\ +nas \end{bmatrix} \_\_\_ \;.\; C$$

Voicing of voiceless coda obstruents before a voiced obstruent

$$\begin{bmatrix} C \\ -voice \\ +obst \end{bmatrix} \rightarrow [+\text{voice}] \;/\; \_ \;.\; \begin{bmatrix} C \\ +voice \\ +obst \end{bmatrix}$$

Devoicing of voiced coda obstruents before unvoiced obstruents

$$\begin{bmatrix} C \\ +voice \\ +obst \end{bmatrix} \rightarrow \begin{bmatrix} devoiced \\ -voice \end{bmatrix} \;/\; \_\_ \;.\; \begin{bmatrix} C \\ -voice \\ +obst \end{bmatrix}$$

Coalescence of coda voiced or voiceless obstruents with their voiceless or voiced counterparts

$$\begin{bmatrix} C \\ +obst \\ \alpha lab \\ \beta dent \\ \varepsilon vel \\ \alpha voice \end{bmatrix} \rightarrow -\alpha voice \;/\; \_\_ \;.\; \begin{bmatrix} C \\ +obst \\ \alpha lab \\ \beta dent \\ \varepsilon vel \\ \_\alpha voice \end{bmatrix}$$

Reduction of double C’s

$$\begin{bmatrix} C \\ \alpha cont \\ \beta obst \\ \varepsilon lab \\ \phi dent \\ \gamma vel \\ \gamma voice \\ \xi nas \end{bmatrix} \rightarrow \;/\; \_\_ \;.\; \begin{bmatrix} C \\ \alpha cont \\ \beta obst \\ \varepsilon lab \\ \phi dent \\ \gamma vel \\ \gamma voice \\ \xi nas \end{bmatrix}$$

Vocalisation of /ʀ/ in coda position

$$[\text{ʀ}] \rightarrow \text{Vi} \;/\; \text{Vi} \_\_ \;.\; \text{C}$$

Deletion of /l/ in “il”-clitics

$$\begin{bmatrix} /l/ \\ +clitic \end{bmatrix} \rightarrow \varnothing \;/\; \_\_\_ \;.\; \text{C}$$

Deletion of voiced C’s after “j’” [ʒ]- clitic

$$\begin{bmatrix} C \\ +voice \end{bmatrix} \rightarrow \varnothing \;/\; [j'] \_\_\_\_\_\_$$

REFERENCES

[1] B. Lindblom, “Explaining phonetic variation: a sketch of the H and H theory”. In *Speech production and speech modelling* (W. Hardcastle and A. Marchal, editors), **Vol 55**, pp. 403-439. NATO ASI Series. Dordrecht, Boston and London: Kluwer Academic Publishers, 1990.

[2]B. Lindblom “Spectrographic study of vowel reduction”, *JASA*, **35**, 1773-1781, 1963

[3]D. Duez,”Consonant sequences in Spontaneous French speech”, *SpoSS*, ESCA, La Baume les Aix, pp.24-26, 1998.

[4]G. Fant, ”Speech sounds and features”. Cambridge, Ma. and London, G.B: The MIT Press, 1973.

[5]O. Fujimura, “Analysis of nasal consonants”, *JASA*, **34**, 1865-1875, 1962.

[6]D. Duez, “On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced plosives”, *Journal of Phonetics*, 25. 407-427, 1995.

[7]C.P. Browman, and L. Goldstein, “Articulatory phonology: An overview”, *Phonetica*, **49**, 155-180, 1992.